

# Supplementary Material

## Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code

Daewoo Pak<sup>1</sup>, Yunsoo Kim<sup>2</sup> and Zachary F Burton<sup>\*3</sup>

\*Corresponding author: Department of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI 48824-1319, Phone: 517-353-0859, Fax: 517-353-9334, [burton@msu.edu](mailto:burton@msu.edu)

1. Center for Statistical Training and Consulting, Michigan State University, E. Lansing, MI 48824, USA
2. Troy High School, Troy, MI
3. Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Rd, E. Lansing, MI 48824-1319; [burton@msu.edu](mailto:burton@msu.edu); 517-353-0859

Supplementary File 1. The *Pyrococcus furiosus* Pfu GlyRS-IIA homology model (PDB file).

Supplementary File 2. Supplementary Figure S1. A sequence alignment of archaeal GlyRS-IIA, IleRS-IA and ValRS-IA enzymes.

Supplementary Figure S1. Alignment of GlyRS-IIA, IleRS-IA and ValRS-IA enzymes. Red shading indicates identity comparing class I and class II aaRS. Yellow shading indicates similarity comparing class I and class II aaRS. Zn-binding motifs are highlighted using green shading. Three  $\beta$ -sheets (sss) that flank the shared Zn-binding domain in class II aaRS are shaded magenta. Shared Zn-finger  $\beta$ -sheets are shaded gray. Active site  $\beta$ -sheets are highlighted in cyan. Numbers indicate e-values for the best local alignments to *Pyrobaculum aerophilum* GlyRS-IIA. HIGH and KMSKS motifs are conserved for class I aaRS enzymes (annotations below the alignment). Motifs 1-3 are conserved for class II aaRS enzymes (annotations above the alignment). Helix is designated hhhh. Not all secondary structures are shown. Pfu) *Pyrococcus furiosus*; Sma) *Staphylothermus marinus*; Sso) *Sulfolobus solfataricus*; Pae) *Pyrobaculum aerophilum*.

				e-value
WP_011012774	-	-----	-----	Pfu GlyRS-IIA 4e-146
WP_011839420	-	-----	-----	Sma GlyRS-IIA 5e-146
WP_009988726	-	-----	-----	Sso GlyRS-IIA 6e-136
WP_011838370	1	M[5] LKQYDPHKVEEW VKRFWDENQIYkLVKEKSDRSILRFNFIDGPPYPSGdVPHIGTAWNKTlkDIVLRYKMR	79	Sma IleRS-IA 0.002
WP_011011404	1	M LPKNYDPNEIEP-[4] ---YWLEEKIY-KYKLEPDKPS--YAIDTPPPFTSG-TLHLGHVLSHTWIDIARYKMR	70	Pfu ValRS-IA 0.86
WP_011839031	1	M SEQRFKPLIEEKR[8] LYKKWEEMGLF-KFTFREDDPREIIAIDTPPPYASG-KWHVGGAAHYAQIDMIARYFRMK	80	Sma ValRS-IA 2.0
Q8ZWU4	1	M[6] LPPSYRPHEVKKA VEEFWRNRRIFeKWK--SWRGGPKFTFLEGPPTNG-MPHVGHIRGRTYKDVVIRFYRL	77	Pae IleRS-IA 6.3
WP_081225707	1	-[3] LTGNYDPKRIEEE IISYWEENKIYnKLKDIVSKRREKFLFIDGPPYPSpTPHIGTIWNKVIKDCILRYQRL	76	Sso IleRS-IA 6.4
		ssssssss HIGH		

WP_011012774	-----		
WP_011839420	-----		
WP_009988726	-----		
WP_011838370	80	GYNVFDPRPGYDCHGLPIEVKVEQKLGVKVKREIeEKIGVDKfVNECKKLVFNNIKSLTKW-FKELGVFMDWENPYLTLRD	158
WP_011011404	71	GYNVLFPPQGFNDHGLPTELKVEKEFGISKDQ-----PEEFLKCKIEWTWQAIEKMRAQ-FIRIGYSADWDLEYHTMDD	142
WP_011839031	81	GYNVLVPFYADRNLGPEVQVEKTYGINPHEISkTPEGREEFLRLCRKFLDKAEEDIVKI-WRMLGCSFDY-WRNGTDSE	158
Q8ZWU4	78	GYDVVWQGGWDMQMPVEWEVEKLLKLRSKKDI-EQFGLEKFALECNLSVEEYLAYWREWgTKRGLGLWLDLENAYETROP	156
WP_081225707	77	GKRVDQPGYDTHGLPIEVATEKLLGI SNKQEIiDKIGVETfINKCKEFALSADKMTQN-FKNVGVFMDWERPYTLDP	155
		ssssss ss	

WP_011012774	-----		
WP_011839420	-----		
WP_009988726	-----		
WP_011838370	159	EYIEAGWWLIKAAEQGLLDREERVVYWCPRCSTTLAEYEV--EYKVLTDPSIYVKFPVRSReKEYLLIWTTPWTLPAN	236
WP_011011404	143	WYKAAVQKSLLDfYKGLIYREEHPVYWCPRCRTSLAKAEV--GYVEEEGYLYYIKLPLADG-SGYIPIATTRPELMPAC	219
WP_011839031	159	EYRKITQATFIELYKGLIYEAERPvNWCPRCKTTLADAEL--EHKEAEGYLYYIKFKVKET-GEEVIVATTRPEMLRAC	235
Q8ZWU4	157	HYLQYAWRIVKRAHELGLLTEDYRVLWfCPRCETSLSDHEValGYDEREDPSIYVKFRVEGGvDEYLVIWTTPWTIVDN	236
WP_081225707	156	SYISSWSVIKKAYEKGMldKGTAVLHWfCPRCETTLSDYEVs-EYRDLEDPSIYVKFKIKGEKNRYLLIWTTPWTIPSN	234
		CXXC	

WP_011012774	-----		
WP_011839420	-----		
WP_009988726	-----		
WP_011838370	237	TFVMAHPD[36] EFPgKKLEGLYDHPLIDIVP[10] VVMAPEFVTTEGTGLVHGAPGHMEDFTVAKKIGIDFIASpI	354
WP_011011404	220	VAVFVHPD DERYKHLVgKKVKLPIYEREV PILADEVDPNFGTGAVYNCTYgDEQDIVWQKRYNLPVIA-I	290
WP_011839031	236	KALVFHPE DDRYKHLNGLHAINPLYGHEM IITTYPEVDPEFGTGLVMVCSYgDQTDVKMFRDLGLKPTII-I	306
Q8ZWU4	237	EAVAVHPD[40] TKKGVELAGVRYIHPLAAEVP[ 8] VVTA-EFVTLEQGTGLVHIAPGHGPEDFELAKKYGLPVTNS-V	354
WP_081225707	235	VFVMINKD[36] TYKGSeligIKYEHPLREFVS[10] VVDAGNIvTLTDGTGLVHSATGHGEEDFTVgQKYGFVVMF-V	351

Motif 1

WP_011012774	1	-----M--KFDKYEYIQDLMRRRGFAWGSFEIYGGARGFYDYGPLGATIKRKIEK-K-IREFIREG-F-FE	61
WP_011839420	1	-----MVNKQVDKYQILLLDLGKRRGIFWLSYEIYGGVAGFYDFGPNVLIKRNININWLNLLVYNIG-LVLE	66
WP_009988726	1	-----MSEVIDK---VLDLAKRRGIFWLSYEIYGGVAGFYDIPVGVRIKKNRIVELMRKYFVKDNGDFVVE	63
WP_011838370	355	DDEGRFTNVAGKYAGKKVREANPEIIKDLKERGALLHASQITHKYPVQWR---CKTPTVMRATKQWVLRVTKLKEKLINE	431
WP_011011404	291	NEDGTMENAGPYAGLKIIEEARKKIAEDLEKMGILYKKEKITHRVLRHTERSSSMPTIELLPKQWFIKPREFTDIVEEV	370
WP_011839031	307	DKDGLRQLVAGPLSGLKVEEARAKAEILEEKGLLVKKEKITKHSVPVQWR---CKTPTQIHTKEWFLKQLDFKDEIRKI	383
Q8ZWU4	355	EINGIYNELGGRYKGVHYDVKETRDLRSGIIVFEEKIRHEYPHCWR---CGSKLILRADRQWFIAISRIIRDKMYAE	431
WP_081225707	352	NDRGEFTEEGKYKGLKVRDASKAII SDLKSKNTLFFEKGLVHRYVQWR---CKTPLILRATDQWFIIRVTIKIKDKMLNE	428

Motif1

WP_011012774	62	IEIPDITPEQVFIASGHVEKQV-----DPLVECKKCGARFRADHIEetLGIDVESKSAEEMT-----KI	121	Pfu GlyRS-IIA	4e-146
WP_011839420	67	IEIPDITPRIVLKASGHEDHFT-----DPITECTRGRVFRADHIEeaLGIRAEGLSMDLW-----KL	126	Sma GlyRS-IIA	5e-146
WP_009988726	64	IESPMIGPAKVFEEASGHVESFT-----DPIVECNNGKRIYRADHIEdiLKKSVVGLKPEELT-----RI	123	Sso GlyRS-IIA	6e-136
WP_011838370	432	AKKVNWIEFDWALDRMMHLENL-QDWWISRQRVWGTPLEIWECEPCHRI-VVGSINE--LEKHG-GKKPKF-----LHRPW	502	Sma IleRS-IA	0.002
WP_011011404	371	AKKINWYPEDMFLRLKDWAESMdwDWWISRQVFGTPLEFWICKNGHII-----PAREED--LPVDP	431	Pfu ValRS-IA	0.86
WP_011839031	384	AKTIEEKPEMHRKLLLDWIDSVSDWPISRDRYATEVFPVWRCRCKGSV-----LLPEPG--kYYRPW	444	Sma ValRS-IA	2.0
Q8ZWU4	432	LQKVNWVETKLRDRFDIFVQNA-RDWNISRVRWGTPLEFVWRCRCKDGRIILVIGSIEE--LKKLAKEIPVDDFklVHRPW	508	Pae IleRS-IA	6.3
WP_081225707	429	IEKVNWIEFDWGKSRISNMVKEL-RDWWISRQRFWGTPLEIWIICRNNVMOVGSKEE--LESIAIDFVEND---LHRPW	501	Sso IleRS-IA	6.4

Motif2

WP_011012774	122	IREHNI--RCPECGGE---LSDV-WYENLMFETYIGPYKDKK-----AYLRPETAQGIIVNEKRLN	176
WP_011839420	127	IKENNI--RCPECGGE---LTKP-KPALLLFKTEIGPYKGGSP-----AYLRPETAQGMFVVFKEHVL	181
WP_009988726	124	IKENNI--RCQYCGGE---LGEV-RLFNLLFTTNGPYTGNI-----GYIRPETAQGMFTAFAKRVY	178
WP_011838370	503	IDEVEI--KCPICGRP---MGRVDPVMDVWEDSAISFVAAN--GHPEKLRLE-EDV-ILDVILEGHDIQIRGWFFSLLRAG	572
WP_011011404	432	FDKPPVekKCPVCCAE---IEPVTDVLDCAVWSSITELIITK [6] GDEEAKKwfEHNFPALRPQGTDIRTWAFYTFIRT	511
WP_011839031	445	KDEPPVdKCPVCCAPkeyLEGEKVFDTWFDSSISVLYVT--GYKRNDKlFRAFKYTLRPQGDIIIRTWLYYSLLRV	520
Q8ZWU4	509	IDQVKI--SAHDCDEW---V-REPYVMDVWLDSGLAWIAAVD GENNRDLW-EKLFYDFVTEGIDQTRGWYFSLATS	579
WP_081225707	502	IDNVRV--KCNKCGV---AKRI PDVADVWFDSSVAFASL- GKDWQEKW-KELGPVDLVLEGHDIQIRGWFFSLLRSG	572

Motif2

WP_011012774	177	AFARNKLPFGVF-----QIGKAYRNEISE----RQGMIRLREFTQAEVE-----	216
WP_011839420	182	NIARNKLPFGIA-----QIGRVGRNEISE----RQGLLRDLREFTIMELE-----	221
WP_009988726	179	EATROLRFIGIA-----QIGKLVGRNEISE----RQGLIRMLREFTIMEIE-----	218
WP_011838370	573	VLGFSKSYNTVLVHGFALDEHGRERHMKSLGNYVGTDEAIEAGRDLRFVWSQNTV-WEDLRFVSRGIDEIRRDLGIAW	651
WP_011011404	512	YMLTGEKEWINDIVINGMVAGPDGRKMSKSYGNVVSFEVPIPKYGADALRLWTALAPP-GEDHPFKWEIVDYNFRFLQKLW	590
WP_011839031	521	YQLTGPAPFKWVRTGMGLDEKGEAMKSKGNVDFEPIYIEKYGADAFRYWAAASAK1GYDYRFBQVIRTGFLFATKWLW	600
Q8ZWU4	580	VLYTGRAPYKNVLIQGLILDKHGQRMKSKGNVWAKDLFEKYGADPVRLYLILSKVApWEDLSFDPDEVKYVIGDLNILW	659
WP_081225707	573	LILLDRAPYTSVLVHGFMLDEQGREHMKSLGNYVEFSVVEKYSRDILRLWLLRNNTT-WEDAKFSWKALELTKRDLQIIW	651

KMSKS

```

-----
----- sss
WP_011012774 217 ----- IFFNPNE--TEHPHE--DEVKDEVLRLY--PIEQKLDIGMI 252
WP_011839420 222 ----- FFFDPEKAFEEETMKYMTDEIRNEKLNII--TAEDKLRGEEKT 261
WP_009988726 219 ----- FFDPEE--NNIENIPLYRFKDTKLNII--TASEKEKD-GKP 255
WP_011838370 652 NTFVFASTYMNldkYDPLQHRIDDYKEF [ 4] DKWLLSRINSIAKKTITSLSEKYYIHEAARELRKFTVEDVSHWYIRLIRP 732
WP_011011404 591 NIYRFAERHIKD--FDYEKYKHIELEPL DRWILSR LHRIKFAFTELEKYRFLNLTRELMTFWHEVADDYIEMIKH 665
WP_011839031 601 NIARF----ISM--FQKPEKKPM-LRPL DEATAYENMMLKSIDFAYSNLDVFPITSYIQFHWNYFASHYIEIVKS 670
Q8ZWU4 660 NVVKFADTYMSLdgFDAEKYPLSQWLEK [ 4] DKWILSELNIMISLFTNFVKNEEFHKAALWREFVVEILSHRYIRLIR 740
WP_081225707 652 NTFVFASTYMNldnFEPDKYTLDDI IKY [ 4] DLWILSRFNSMLKKNVNSMKDYKVHEMTNYLINFIEDVSRFYIRLIRK 732

```

```

----- sssss ----- sssssss
WP_011012774 253 EL-TAEHA-----VFKGYVMNTFFAYYVMVIKRIILLDIGIPEDKIRFRQOLPE-----ERAHYSADTWDA 311
WP_011839420 262 RKYTAIRL-----ENKIVKNPWMAYWMALGNIFLRKLGISSEKIRFDEKLPE-----EKAHYSQTFDQ 321
WP_009988726 256 QEFKLEES-----VREKIIIIQPWMAYWMAVASSFVKALGINE--FYEEKLPY-----ERAHYSKQTFDQ 313
WP_011838370 733 RVVVEENT---PDKLAAYSVLYYVLEKMLRMMAPFPFPLAEKIYQEVFRKANPElpLSIHLLDWP SIDDEYIDELEKIM 809
WP_011011404 666 RLYGEDL---ESKLRKAKVALYELLYNIMLLLAPFVPHITEELYHHIFKEKIGI--KSVHLLQWPEYREDRIDEAEKIG 739
WP_011839031 671 RAYNRENKysrEEQEAAWYTLHYILNRTLRLALAPMPFVTDALYTRYLG-----KTVHLEKFPPEPEKEWLNKDYQLAL 743
Q8ZWU4 741 RVWSEEPS---PDKYAAAYAVLHDVLKQVLIILGSLIPFITEYLWQAVVRKYEKNapESVHQAQYPAAGS--YDKLIIYAY 815
WP_081225707 733 RAWIEANT---QDKLAMYVILLYILKQVLIILASTIIPFISEKIYKSFVVNAKE----SVSMESSINYDERFIDNLERAF 805

```

Motif3

```

s sssssssss
WP_011012774 312 EIYSRRGHW-VECVGLAYRTDYDLSRHMKMSGADLTVMIHYDKpKIIVKLNVTLNMKV [5] EKV-KEINEKLSMSQEE 391 Pfu GlyRS-IIA 4e-146
WP_011839420 322 EVYTKYGW-IEVAGYSYRTTYDLSRHIKYSKADLTYPFKKYDKpIKKIVKIVPAPKKI [1] EIVGKDIGTVMKKLNEIF 398 Sma GlyRS-IIA 5e-146
WP_009988726 314 --LAVINGIKVEISGHAYRGDYDLSRHMKYSGQDLTIFRKYKpKMIKKKTVIIVNNTKL [3] ELR-KKIMELLNGKSAEE 390 Sso GlyRS-IIA 6e-136
WP_011838370 810 NVIREIYEAAAAARMKAGIKLRQPVK-SLTVYTRERIREVTR-KYSGLLARLVNVRKV EAKQVAEIGKIVKYKVSF 884 Sma IleRS-IA 0.002
WP_011011404 740 ELAREIVSAMRKYKNSHGMPLNAKLKHVAIYATSYEMLKVIE-KD---IAGTMNIERL EIVKGEPPQLEEKVVEIKP 812 Pfu ValRS-IA 0.86
WP_011839031 744 -LVQKINSAIWSYKKKNNIKFKEKLT-KKIYLP IEAEKAL---KE---IVDLHKLVNI EFHSGSPPA----- 802 Sma ValRS-IA 2.0
Q8ZWU4 816 RELFAVFSALAEARNKAGIKLRWPIR-EAYV-----NGAKYAE-RYTELLKYLGNVKEV KVGR----- 871 Pae IleRS-IA 6.3
WP_081225707 806 EVAREINEASLNARAKAGIKLRWPLA-KVYIFIE NEDTLAKVG-RIKDVLISMLNAKDI EISKIEGFKSFSKYKVEF 880 Sso IleRS-IA 6.4

```

```

----- sssssssss
WP_011012774 392 LRKLVEEINAN GKVIIEGHELEKEDI I I KEVEEKIQGEKIV-----PHVLEPSFGIDRPFYLLLENSLTVD 459
WP_011839420 399 AEKILEELNKN NYVEISNYKLSRECFIISEKETEIHGEKII-----PHVVEPSFGLERIFVVLLENSISTARD 466
WP_009988726 391 IEQMIR-----NNVMIDDKPLS-EFINIVERREKVVHGEHFI-----PHVVEPSFVERTLFLTILSAYK-EKE 451
WP_011838370 885 IYRVIGPLYRK [6] NYIMENQES IARDIIRKGEHTAYIEGQETKLTKEQVLITPSYV-EGYSVEREWGVSVAIDTRLSKE 966
WP_011011404 813 IYKRI GPRYK [6] KHLQENAEIIGRIKEEGKVEFEVEGQKVLEKEDIEIKKAVFSEGEEVET----- 880
WP_011839031 803 -----NAEKIG-----EEVYIEK----- 815
Q8ZWU4 872 -----RPDYLCVKEGEL-E-----VCVPDKTEPE 894
WP_081225707 881 NRSIIGKEYKS [6] EYIENNRDI IAMDILNKKQHVAKIDNFDIILNASYV I I SEETV-EGFISSKFSKGIVVIKETSSES 962

```

WP_011012774	460	G	R	V	Y	L	K	I	K	D	M	A	P	E	V	A	V	L	P	L	M	A	----	K	P	T	T	I	A	E	I	Y	R	N	L	Q	K	E	G	F	I	V	V	L	E	K	D	S	I	G	R	Y	---	M	R	Y	D	E	I	G	T	P	Y	C	527											
WP_011839420	467	G	R	I	V	L	K	L	P	P	R	I	A	F	Y	K	V	A	V	F	P	L	M	T	G	N	K	P	E	H	K	K	M	S	I	A	R	Q	V	Y	R	A	L	L	E	N	N	I	L	A	Y	L	E	E	G	S	I	G	R	Y	---	A	R	A	D	E	I	G	I	P	Y	A	539			
WP_009988726	452	G	R	V	L	L	S	L	P	R	H	L	A	F	Y	D	I	A	V	F	P	L	L	E	----	R	E	E	L	R	K	S	V	E	I	R	E	S	L	S	E	K	-	Y	E	V	L	Y	D	S	G	S	I	G	R	Y	---	A	R	A	D	E	I	G	I	P	F	A	518							
WP_011838370	967	E [11]	R	I	Q	V	M	R	K	E	I	N	L	E	L	D	A	K	I	E	T	Y	I	A	P	R	K	H	V	E	-	L	L	K	P	Y	T	E	Y	I	K	N	E	T	R	S	E	K	L	T	Y	L	L	T	P	E	Q	L	E	K	I	Q	g	y	r	K	E	W	E	I	Q	E	Q	Y	I	1052
WP_011011404		-	-----																																																																									
WP_011839031		-	-----																																																																									
Q8ZWU4	895	L [11]	R	I	Q	V	M	R	K	E	T	G	L	E	I	S	D	E	I	H	V	V	E	T	N	S	D	D	I	K	K	A	M	E	Q	Y	R	D	Y	I	A	R	E	T	R	A	V	K	L	I	I	----	D	A	V	S	Q	----	g	K	E	W	D	I	S	E	K	V	K	973						
WP_081225707	963	E [11]	R	I	Q	F	M	R	K	Q	L	K	L	N	V	L	D	Y	I	E	I	S	M	K	V	P	E	E	R	V	K	-	T	I	Q	K	W	E	E	F	I	K	S	E	T	R	A	S	N	I	I	L	G	E	A	K	G	D	I	T	M	----	D	W	D	I	E	G	E	S	Y	I	1042			

WP_011012774	528	V	T	V	D	N	Q	L	P	E	D	G	[30]	568
WP_011839420	540	I	T	I	D	Y	Q	L	E	D	E	[38]	588	
WP_009988726	519	I	T	V	D	P	Q	L	V	D	N	[45]	574	
WP_011838370	1053	I	V	I	K	Q	L	1058						
WP_011011404	881	A	V	R	D	A	T	I	L	F	F	891		
WP_011839031		-----												
Q8ZWU4	974	I	E	R	K	A	Q	A	981					
WP_081225707	1043	I	G	I	K	K	S	T	1049					